Complex Wavelet Based Envelope Analysis for Analytic Spectro-Temporal Signal Processing

Luneau, Jean-Marc

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Luneau, J-M. (2008). *Complex Wavelet Based Envelope Analysis for Analytic Spectro-Temporal Signal Processing*. Aalborg Universitet.

Complex Wavelet Based Envelope Analysis for Analytic Spectro-Temporal Signal Processing

Jean-Marc LUNEAU[≀],
Dr. **Jérôme LEBRUN**[†]
& Prof. **Søren Holdt JENSEN**[≀]

February 2008

[≀]Aalborg University

MISP
Department of Electronic Systems
Niels Jernes Vej 12, A6
DK - 9220 Aalborg Ø, DENMARK

[†]Centre National de la Recherche Scientifique

Laboratoire I3S
Algorithmes/Euclide B
2000 route des Lucioles - BP 121
F - 06903 Sophia Antipolis, FRANCE

AALBORG UNIVERSITY

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Complex wavelet based envelope analysis for analytic spectro-temporal signal processing

Jean-Marc Luneau, Jérôme Lebrun and Søren Holdt Jensen

February 2008

Abstract

A new class of spectro-temporal signal transforms often called Modulation Transforms has recently been introduced. They add a new dimension to the classical time/frequency representations, namely the modulation frequency. Very efficient in many applications, especially at the analysis level, these transforms show however their limits when it comes to signal processing in the transform domain. The modulation spectrum is commonly obtained by spectral analysis of the sole temporal envelope of the sub-band signals. Simple filtering in this domain has been shown to create serious distortions. We detail here some of the reasons for this and propose the use of a complex wavelet transform as a more appropriate envelope and phase processing tool. By working in an alternative transform domain coined as *Modulation Sub-bands*, this transform shows very encouraging denoising capabilities and suggests new approaches for joint spectro-temporal analytic signal processing in general.

This work was supported by the European Union via a Marie Curie Fellowship (EST-SIGNAL program http://est-signal.i3s.unice.fr) under contract No MEST-CT-2005-021175.)

Introduction

Many natural signals can be seen as the sum of low frequency modulators of higher frequency carriers. For instance the concept of modulation frequency [1, 2, 3, 4] appeared to be very useful to analyze speech or broadband acoustic signals in general. The focus of this paper is to build a proper modulation frequency sub-band analysis and take advantage of all the information and possibilities the spectro-temporal transform may provide for signal processing. The starting point is to highlight the fact that so far, the phase signal in the modulation sub-bands was either ignored or processed in an *ad-hoc* way with many efficiency/artifacts compromises.

In this paper we first show the relevance and importance of spectro-temporal approaches for audio and speech signals in particular. Then we describe and explain the limitations of the usual implementations. This motivates the need for alternative ways of building the transform. Then to get analyticity, we explain the motive behind using complex wavelets instead of a classical Fourier analysis. We detail our filterbank method and show some interesting outcomes of this transform on envelope and phase analysis first and then on denoising. We finish by a discussion on the results and propose methods for future improvements in order to get Complex Wavelet Modulation Sub-bands decompositions and conclude with its legitimacy for future speech related applications and general spectro-temporal signal processing.

1 Spectro-temporal signal processing

1.1 Concept of modulation spectrum

Recent researches have explored three-dimensional energetic signal representations where the second dimension is the frequency and the third is the transform of the time variability of the signal spectrum. This signal spectrum is a time-acoustic frequency representation, *i.e.* a usual short-time or long-time Fourier decomposition of the signal. The third dimension is usually called the "modulation spectrum[1]" . The second step of this spectro-temporal decomposition can be viewed as the spectral analysis of the temporal envelope in each acoustic frequency band. It provides three-dimensional information from the signal with two-dimensional energy distributions $S_t(\eta, \omega)$ along time t with η being the modulation frequency and ω the acoustic frequency.

[1]The name of Modulation Spectrum may vary from author to author.

1.2 Physiological background for audio signals

In the speech and audio signal processing communities, there has been a strong motivation to thoroughly study the modulation structures since Drullman *et al.* [5], refined later by Greenberg [1], showed that the modulation frequency range of 2-16Hz has a special role for speech intelligibility. It reflects the syllabic temporal structure of speech [1]. It is also demonstrated that the modulation frequencies around 4 Hz are the most important for human speech perception. In general, low frequency modulations of sound seem to carry significant information for speech and audio. The modulation spectrum is thus particularly well fitted for speech signals. But it is also a very delicate representation to work on [6]. Any artifact introduced in a speech signal is obvious and changes very easily its "natural" sounding. This is a call for a better understanding of the modulation structures in speech but also for alternative representations in which the perceptually important spectro-temporal information should be better decorrelated so as to allow efficient and non-destructive processing in that domain.

For acoustical signal processing in general there are more important facts to take into account. The first aspect is the signal phase too often ignored when it comes to digital audio processing: two signals with identical magnitude spectra but different phases do sound different. Ohm's acoustic law stating that human hearing is insensitive to phase is persistent but wrong. For instance, Lindemann and Kates showed in 1999 [7] that the phase relationships between clusters of sinusoids in a critical band affect its amplitude envelope and most important, affect the firing rate of the inner hair cells. Thus the major issue is to preserve the phase during a modulation transform otherwise amplitude envelopes will be modified. Magnitude in a signal gives information about the power while phase is important for localization. For the human hearing, studies like [8] showed that the basilar membrane in the cochlea, basically acts like a weighted map that decomposes, filters and transmits the signal to the inner hair cells. If the phase is altered the mapping on the membrane may be slightly shifted hence the different sounding.

The second important fact to take into account for digital audio and speech processing is the mechanical role of the human hearing system and particularly the middle ear and the cochlea. Different studies [6] showed that for frequencies below approximately a threshold of 1.5-2kHz (and gradually up to 6kHz) the firing rate of the inner hair cells depends on the frequency (and on the amplitude and duration) of the stimulus. At those frequencies it is called time-locked activity or phase locking, *i.e.* there is a synchrony between the tone frequency and the auditory nerve response that becomes progressively blurred over this threshold. From 2kHz and above 6kHz, the response of the inner hair cells is function of the stimulus signal envelope and the phase is less important [9].

1.3 Applications

Multiple topics have been investigated with relative success over the past few years with this modulation transform: pattern classification and recognition [3], content identification, signal reconstruction, watermarking, single channel source separation, audio compression, automatic speech recognition, just to name few. It was also experimented in the area of speech enhancement (pre-processing method) for improving intelligibility in reverberant environments [10, 11] or speech denoising [12] but there again with some limitations. The experiments had to face either a production of severe artifacts or a recourse to post-processing in order to get rid of musical noise.

1.4 Analytic modulation spectrum

The classical computation of the modulation spectrum relies on the envelope detection of analytic signals or quadrature representations obtained using the Hilbert transform in each sub-band. More precisely, the input signal $x[n]$ is decomposed into M sub-band signals $X_k[n]$ using typically a bank of modulated filters $h_k[n]$ where $k = 0, \dots, M-1$ is the sub-band index. When using real filters as in the ubiquitous MDCT, the extraction of the envelope in each sub-band is usually done with the Hilbert transform $H\{.\}$ by introducing $\tilde{X}_k[n] := X_k[n] + jH\{X_k[n]\}$, *i.e.* an analytical extension of $X_k[n]$. Each sub-band signal can be decomposed into its envelope

$$A_k[n] := |\tilde{X}_k[n]| \tag{1}$$

and its instantaneous phase

$$p_k[n] := \cos(\varphi_k[n]) \tag{2}$$

with $\tilde{X}_k[n] = A_k[n].e^{j\varphi_k[n]}$. We then get for real-valued sub-bands, $X_k[n] = A_k[n]p_k[n]$. The modulation spectrum in the k^{th} sub-band is then obtained by computing the Fourier transform of the envelope signal $A_k[n]$.

With this approach, any filtering or processing of the sub-bands introduces artifacts and distortions at the reconstruction. As stressed in [13], this is essentially due to the way the envelope signal is obtained. Also, using, as described in [3], a real-valued wavelet transform on the envelope signals $A_k[n]$ instead of the classical Fourier transform doesn't bring much improvement since the core issue is really the envelope extraction. Namely, processing the modulation spectra without taking much care of the phase signals p_k leads to a leakage of energy from the modified sub-band to other sub-bands. The reconstructed sub-band being the product in time-domain of the modified envelope and the original carrier (giving a convolution in the Fourier domain), the bandwidth of the modified sub-band may then be wider than the original one. This will lead to imperfect alias cancellation between the sub-bands and thus artifacts. Phase degradations imply a bad envelope reconstruction and severe artifacts that are very noticeable in speech signals.

An *adaptative* way to circumvent this problem for modulation filtering was proposed by Schimmel and Atlas [13]. In order to reconstruct sub-bands achieving narrow bandwidth and thus little leakage and artifacts, they suggested the use before-hand of a "coherent" carrier detection to get a $\tilde{\varphi}_k[n]$ signal closer to the true phase of the signal but also narrowband. It implies that both the envelope and the carrier are now complex, so equations (1) and (2) become:

$$A_k^c[n] := \tilde{X}_k[n].e^{(-j\tilde{\varphi_k}[n])} \tag{3}$$

and

$$p_k^c[n] = e^{j\tilde{\varphi_k}[n]} \tag{4}$$

where $\tilde{\varphi}_k[n]$ is a low-pass filtered version of the estimated phase signal. Their idea is to design this low-pass filter by compromising the desired amount of distortion and the effectiveness of modulation filters stop-band attenuation. However, by its highly adaptative and non-linear nature, this approach makes it again quite difficult to do any processing in the transform domain.

1.5 Necessity of a new approach

We introduce here an alternative method that simply avoids the computation of the envelope signals but nevertheless provides a time-scale version of the modulation spectrum for each sub-band. The underlying idea in our approach is motivated by the fact that extracting the envelope is "similar" to extracting the polynomial part of the signal. And this is a typical job in which wavelet transforms do well. Hence, the principle is now to perform wavelet transforms on each sub-band to extract the polynomial parts. More precisely we will use complex-valued wavelets as they allow us to deal properly with the phase in the signals. The goal is to achieve a decomposition of each subband signal into different complex-valued modulated components living each into disjoint sub-spaces tiling the modulation frequency dimension. Strictly speaking, this spectro-temporal approach cannot be called modulation spectrum anymore as we work with polynomial approximations coming from the wavelet processing. This decomposition should be called a *Modulation subspaces* decomposition instead. With this new approach, many improvements should be possible not only in speech or audio enhancement but also in spectro-temporal related application domains in general.

2 Complex wavelet method

The problem with most spectro-temporal or modulation frequency frameworks is often the lack of resolution at the crucial low modulation frequencies. This drawback comes again from using the Fourier analysis as second transform in the process as it only permits a uniform frequency decomposition which yields to uniform modulation frequency resolution. A log frequency scale allows to adapt the precision on the important modulation frequencies between 2 and 16 Hz [3]. Moreover, from a psychoacoustic point of view [14], such a scale matches better the human perceptual model of modulation frequencies, hence again the idea of using a wavelet transform as second step of the modulation transform, especially for natural or speech signals.

2.1 complex vs. real valued wavelets

The discrete wavelet transform has been a successful new tool in many fields of signal processing and especially in image processing. In brief, the idea underlying wavelets is to replace the infinitely oscillating sinusoidal basis functions of Fourier-like transforms by a set of time/scale localized oscillating basis functions obtained by the dilatations and translations of a single analysis function, the *wavelet*. Nevertheless, with the first generation of real-valued wavelets, it was difficult to deal properly with both amplitude and phase informations in a signal. This explains partly their limited success in audio and speech processing. However, the recent developments of new complex-valued wavelet-based transforms [15] alleviate most of these limitations. Complex wavelets have the property to deal properly with both amplitude and phase of the signal which is a crucial matter as seen earlier in section 1.2.

2.2 Filterbank specifications

It has been shown that by using complex wavelets, one can implement new filterbank structures that ensure the analyticity of the analysis [16]. As usual, the filterbank will be used in an iterated manner. Indeed, the analysis of a signal at several scales (multi-resolution analysis) can be accomplished by iterating the filterbank on the low-pass sub-band. The idea of filterbank trees is to cascade this iteration up to a certain level l. We then have $l + 1$ signals: a coarse version and l detail signals. The original signal can then be reconstructed from these sub-band signals by iterating on the synthesis filterbank with the following specifications:

Signal phase As seen before, the focus of our work around modulation transforms is on the signal phase information. The filterbank has to deal properly with it in order to cautiously characterize the energy localization of the signal.

Polynomials Working with wavelets means we are not anymore in the frequency domain, the result is not a spectrum. The transform has to process and preserve polynomials from scale to scale of the multi-resolution wavelet decomposition.

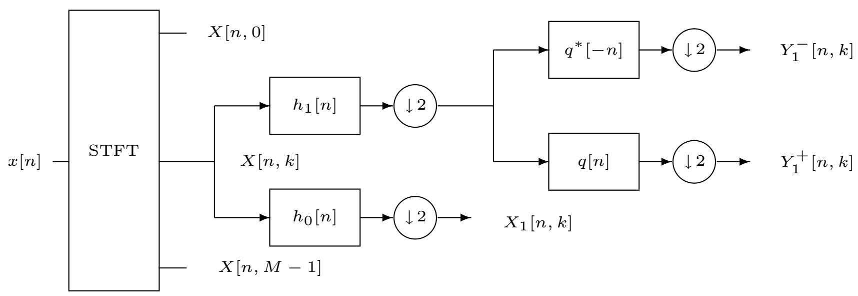
Analytic Hilbert-like transform An analytic transform is needed but wavelets are not appropriate to compose a Hilbert transform so the goal is to approximate it. Van Spaendonck *et al.* showed in [16] that two separate high-pass complex filters are able to distinguish high positive frequencies from high negative ones. Hence, the real and imaginary parts of the resulting complex wavelets would be similar to Hilbert transform pairs.

Redundancy vs. Orthogonality If we want the results to be well decorrelated to do some sub-band filtering or any thresholding it is easier if the filterbank is orthogonal. However, in 2001, Selesnick [17] proposed a shift-invariant complex wavelet transform where the scaling and wavelet functions form an Hilbert transform pairs. This so called dual-tree complex wavelet transform has a major drawback to be redundant but provides a tight frame structure with redundancy two that makes it at least as efficient as an orthogonal framework. We first aim here at a unidimensional filterbank but research with redundant dual-tree transforms will most surely be investigated in the future to see if they could bring any further improvement.

FIR Filters must have a finite impulse response to guarantee good time localisation.

2.3 Complex Wavelet Modulation Sub-Bands

Here, by working with complex wavelets, we avoid the limitations of the usual Hilbert envelope approaches caused by the separate processing on the magnitude and the phase of the modulation spectrum in the sub-bands. In our approach the sub-band signals $X_k[n] = X[n,k]$ are obtained using a complex modulated filter-bank (a Short Time Fourier Transform here) for $k = 0, \ldots, M-1$ and further decomposed using an orthogonal complex wavelet filterbank as shown in



where $X_1[n,k]$ is a coarse version of the sub-band signal $X[n,k]$ and $Y_1^+[n,k]$ and $Y_1^-[n,k]$ are respectively the positive and negative frequency components of the associated detail signal. The complex wavelet filterbank is then iterated N times on each lowpass signal obtained $X_1[n,k], X_2[n,k], \ldots$. Here, motivated by their good phase behavior, we took $h_0[n]$, $h_1[n]$, $g_0[n]$ and $g_1[n]$ to be orthoconjugate complex Daubechies wavelet filters [18]. More precisely, we did our experiments using the complex Daubechies filters of length 10 based on the the low-pass filter $g_0[n]$ given in Table 1.

Now, $q[n]$ is a bandpass orthogonal filter that satisfies the conditions given in [16] to get analyticity, *i.e.* it is obtained from a complex-valued lowpass orthogonal filter $u[n]$ satisfying

$$U^*(1/z)U(z) + U^*(-1/z)U(-z) = 2.$$

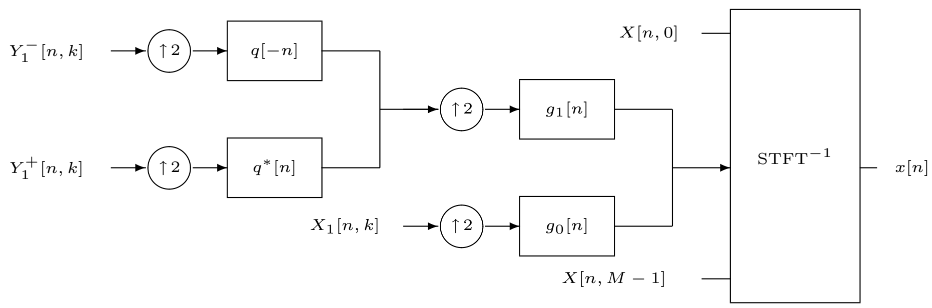
In our case we took $q[n] := j^n u[n]$ where

$$u[n] = \frac{\sqrt{3}}{16}[-1,0,5,5,0,-1] + j\frac{\sqrt{5}}{16}[0,1,3,3,1,0]. \tag{5}$$

Table 1: Coefficients for orthoconjugate complex Daubechies filters of length 10

n	$g_0[n]$	
0	0.01049245051230	+0.02059043708702j
1	-0.00872852869034	-0.01712890812780j
2	0.08063970414533	+0.11794747353812j
3	-0.09422365674476	-0.15137970843150j
4	0.64300323451588	+0.18285216450551j
5	-0.18285216450551	+0.64300323451588j
6	-0.15137970843150	+0.09422365674476j
7	-0.11794747353812	+0.08063970414533j
8	-0.01712890812780	+0.00872852869034j
9	-0.02059043708702	+0.01049245051230j

The reconstruction is then done using the complementary synthesis filter-bank.



Now, we omit some details of the signals at the reconstruction by picking only the *relevant* coefficients in the decomposition - this is the underlying principle of denoising by sparse representations [19, 20]. Indeed, for a *well designed* reconstruction basis, the noise is not picked in the sparse coefficients used to reconstruct the signal, hence the denoising. The "quality" of the reconstructed signal depends largely on the choice of the basis vectors with which the reconstruction is performed. In our case, the dual stage synthesis, inverse Complex-DWT followed by inverse STFT, gives reconstruction vectors that are well adapted to acoustical signal processing, as proved below, namely dilated windowed sinusoidal functions similar to scaled Gabor functions.

Furthermore, this decomposition separates the complex-valued components obtained (*i.e.* with proper magnitude and phase) into orthogonal spaces. With this method, if we do any thresholding or remove sub-bands from the wavelet decomposition, we do not create aliasing problems between the sub-bands. Typically, if some uncorrelated noise is spread on the modulation sub-bands, for each of them the phase and the magnitude can be properly cleaned. We are thus insured not to widen the spectral bandwidth of the sub-band and thus not to smear on the near-close sub-bands.

3 Detection and denoising experiments

The purpose of our experiments is to illustrate the relevance of the second part of the transform, *i.e.* the complex wavelet filterbank, as a proper envelope and phase "detector". The test signals are taken to be windowed complex chirps (see Fig. 1). So, the sub-band signals $X_k[n]$ from the first transform are assumed to be sampled versions of the type

$$c(t) = w(t).e^{j2\pi(\omega_1 t^r + \omega_0)}$$

where $w(t)$ is a piecewise polynomial envelope, ω_0, ω_1 frequency parameters and r characterizes the frequency evolution. This is a good model for the sub-bands signals $X_k[n]$ when $x[n]$ is assumed to be voiced speech and the first transform is a STFT as in this paper. Fig. 1 gives the real and imaginary parts of $X_k[n]$ on 2^{18} samples with a quadratic envelope and parabolic phase.

3.1 Detection

The thresholding of the coefficients obtained in the Modulation sub-bands provides approximations of the polynomial parts of the envelope and the phase signals.

When choosing to reconstruct without thresholding with all the sub-bands the reconstruction is perfect. Now, for instance in Fig. 3 if we choose an order of 10, *i.e.* we lowered the resolution by 8 scales at the reconstruction by keeping only the coarse Modulation sub-band $X_8[n, k]$. This represents a simple yet severe thresholding of the higher sub-bands (and compression by a factor of 2^8). Nevertheless, the reconstruction is still almost perfect.

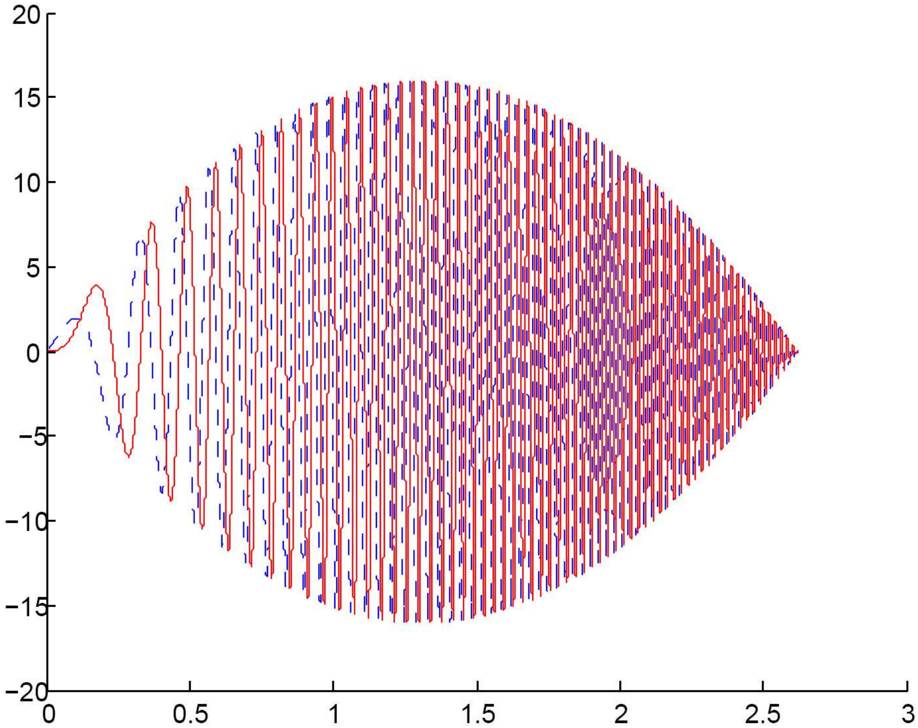


Figure 1: *Input signal $X_k[n]$: real (dashed) and imaginary (solid) parts*

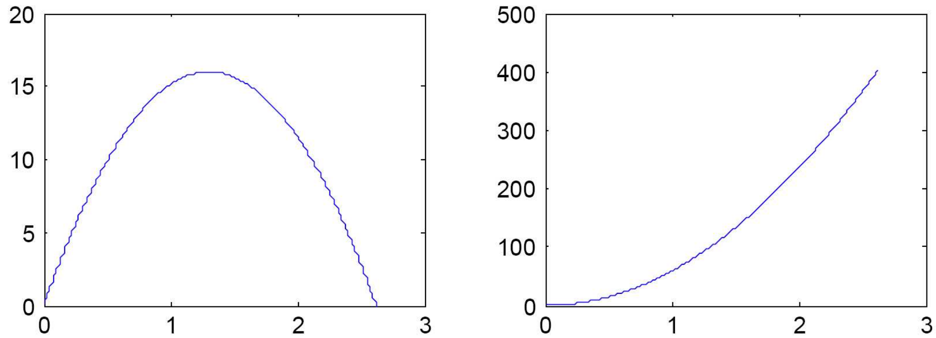


Figure 2: *Original signal: amplitude and phase*

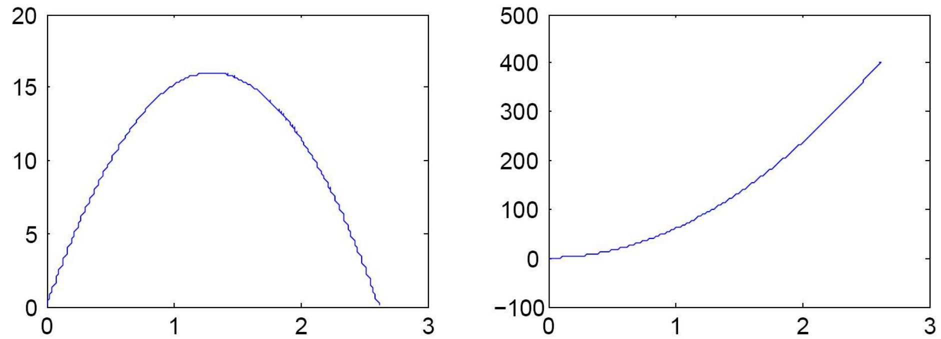


Figure 3: *Reconstructed signal: amplitude and phase*

Amplitude and phase remain almost identical to the original ones. This means our analysis/synthesis structure is well adapted to signals of this kind. This implies good denoising possibilities because the thresholding is altering neither the envelope nor the phase. This is what we wanted so as to work properly on the sub-bands of the spectro-temporal transform without causing artifacts like described earlier in section 1.

3.2 Denoising

To evaluate the denoising capacities of the transform by a simple hard thresholding, we tested the same protocol of reconstruction on the same original signal but now drowned in severe white noise, as shown in Fig. 4. The hard threshold used is of the form $T = \sigma\sqrt{2\log_e N}$ (with σ^2 the noise

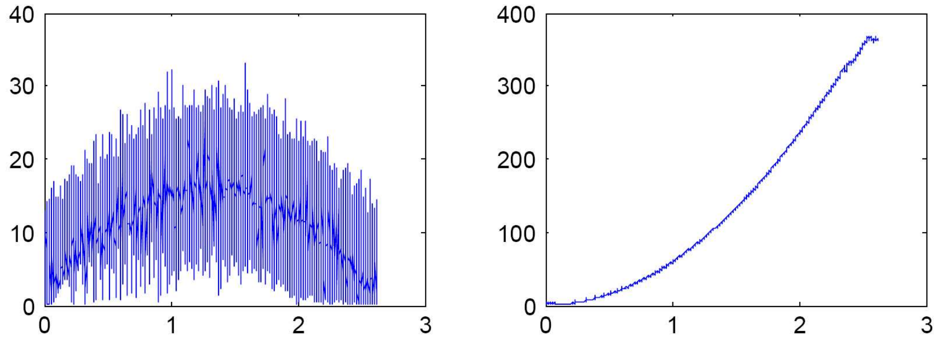


Figure 4: *Noisy signal: amplitude and phase*

variance and N the size of the basis we reconstruct with, [21, 20]).

Results of the reconstruction with thresholding at resolution scale 2^{10} out of 2^{18} are shown in Fig. 5. The envelope has been recovered and most of the noise has been removed, which was expected from wavelet polynomial approximation of piece-wise continuous signals. The novelty concerns the phase which has been also nicely smoothed. This is a major improvement on the wavelet approach described in [3]. For comparison, we applied their approach on the same signal: our noisy sub-band signal is split into its amplitude and phase signals, wavelet thresholding is performed on the sole amplitude signal and at a final stage the original phase is *glued* back on the denoised amplitude signal to reconstruct a denoised version of the sub-band signal. As illustrated in Fig. 6, this typical way of doing signal processing on the modulation spectrum shows its limits: only the outer-envelope of the noisy signal is restored and the phase is completely lost.

As stated in section 2.3, magnitude and phase of the signal are cleaned jointly. Furthermore, for a future spectro-temporal processing we do not want the spectral bandwidth of a sub-band to be widen. Indeed, with this transform, a clean signal is perfectly reconstructed and if the signal undergoes denoising, phase and envelope are smoothed hence the sub-band spectrum can only become narrower.

4 Discussion

With this example, we illustrated some of the properties of this complex wavelet transform. The envelope and the phase are indeed well processed and this second part of the transform provides a nice framework for more

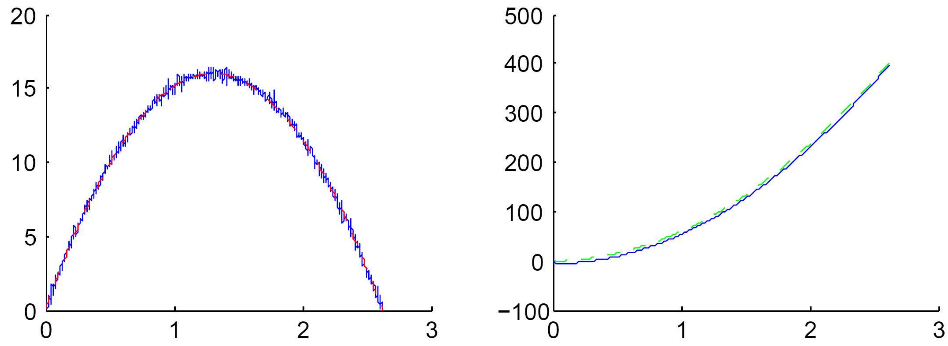


Figure 5: *Denoising results with the complex wavelet transform on the envelope and phase - comparison with the original signal (dashed)*

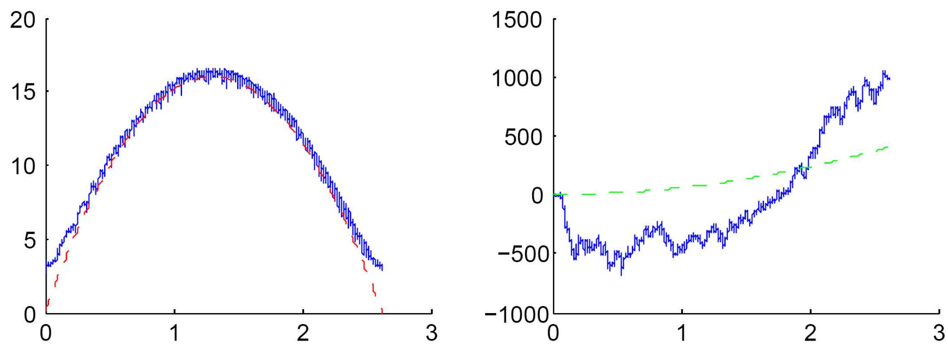


Figure 6: *Denoising results obtained with the usual approach - comparison with the original signal (dashed)*

advanced processing tools to work on the modulation structures. However some work is still to be done on the first part of the transform. Namely, in this paper we used a STFT, which provides a good access to the phase information but also some limitations on the sparsity of the representations we can achieve. Namely, the denoising performances depend on the capacity to approximate the signal with very few basis vectors.

Also, in this paper, to motivate our approach, we used a very coarse method for denoising: skipping some sub-bands with basic thresholding of coefficients. In general, denoising by wavelet thresholding can be done in (quasi-minimax) optimal ways. In our case, a refined multi-scale SURE thresholding, *i.e.* adaptive thresholds in the sub-bands, would be a much more adequate technique. There exist also more advanced algorithms like Orthogonal Matching Pursuit and other sparse representations techniques that gives better results for specialized distortion measures [20].

Conclusion

In this paper we introduced a new way to process modulation frequencies using complex wavelets. We proved the legitimacy of this approach since the transform we proposed is based on complex wavelets which deal properly with both magnitude and phase informations in the signal. We showed that envelope and phase are processed very properly which provides a big margin for denoising. On a rather simple experiment of hard thresholding in the sub-bands, some basic but very efficient denoising has been shown. In a near future, other options for the first part of the full transform (a STFT so far) will be investigated to take better profit of the second part, the complex wavelet filterbank. Also, more sophisticated denoising tools coming from the sparse representations theory will be tried on some real-life examples.

Acknowledgments

This work was supported by the European Union via a Marie Curie Fellowship (EST-SIGNAL program http://est-signal.i3s.unice.fr) under contract No MEST-CT-2005-021175.)

Bibliography

[1] S. Greenberg, "On the origins of speech intelligibility in the real world," *Proc. of ESCA Workshop on robust speech recognition for unknown communication channels*, pp. 22–32, 1997.

[2] Q. Li and L. E. Atlas, "Properties for modulation spectral filtering," *Proceedings of ICASSP '05*, vol. 4, pp. 521–524, Mar 2005.

[3] S. Sukkittanon, L. E. Atlas, and J. W. Pitton, "Modulation-scale analysis for content identification," *IEEE Transactions on Signal Processing*, Oct 2004.

[4] H. Hermansky, "The modulation spectrum in the automatic recognition of speech," *Automatic Speech Recognition and Understanding*, pp. 140–147, Dec 1997.

[5] R. Drullman, J. M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech perception," *JASA*, vol. 95, pp. 1053–64, February 1994.

[6] J. Blauert, "Spatial hearing: The psychophysics of human sound localization," *MIT Press*, 1997.

[7] E. Lindemann and J. M. Kates, "Phase relationships and amplitude envelopes in auditory perception," *Proceedings of WASPAA*, pp. 223–226, 1999.

[8] L. Golipour and S. Gazor, "A biophysical model of the human cochlea for speech stimulus using STFT," *IEEE ISSPIT*, 2005.

[9] D. H. Johnson, "The relationship between spike rate and synchrony of auditory-nerve fibers to single tones," *JASA*, vol. 68 (4), pp. 1115–1122, October 1980.

[10] A. Kusumoto, T. Arai, K. Kinoshita, N. Hodoshima, and N. Vaughan, “Modulation enhancement of speech by a pre-processing algorithm for improving intelligibility in reverberant environments,” *Speech Communication*, vol. 45(2), Feb 2005.

[11] C. Avendano and H. Hermansky, “Study on the dereverberation of speech based on temporal envelope filtering,” *Proceedings of ICSLP 1996*, pp. 889–892, 1996.

[12] H. Hermansky, E. A. Wan, and C. Avendano, “Speech enhancement based on temporal processing,” *Proceedings of ICASSP 1995*, pp. 405–408, May 1995.

[13] S. Schimmel and L. E. Atlas, “Coherent envelope detection for modulation filtering of speech,” *Proceedings of ICASSP ’05*, vol. 1, pp. 221–224, mar 2005.

[14] T. Houtgast, “Frequency selectivity in amplitude-modulation detection,” *JASA*, vol. 85, pp. 1676–80, 1989.

[15] I. W. Selesnick, R. G. Baraniuk, and N. Kingsbury, “The dual-tree complex wavelet transform - a coherent framework for multiscale signal and image processing,” *IEEE Signal Processing Magazine*, vol. 22(6), pp. 123–151, Nov 2005.

[16] R. van Spaendonck, T. Blu, R. Baraniuk, and M. Vetterli, “Orthogonal hilbert transform filter banks and wavelets,” *Proceedings of ICASSP ’03*, vol. 6, pp. 505–508, Apr 2003.

[17] I. W. Selesnick, “Hilbert transform pairs of wavelets bases,” *IEEE Signal Processing Letters*, vol. 8 (6), 2001.

[18] X.-P. Zhang, M.D. Desai, and Y.N. Peng, “Orthogonal complex filter banks and wavelets: some properties and design,” *IEEE Trans. Sig. Proc.*, vol. 47(4), pp. 1039–1048, 1999.

[19] R. Gribonval and E. Bacry, “Harmonic decomposition of audio signals with matching pursuit,” *IEEE Trans. Sig. Proc.*, vol. 51(1), pp. 101–111, 2003.

[20] M.E. Davies and L. Daudet, “Sparse audio representations using the MCLT,” *Signal Processing*, vol. 86, no. 3, pp. 457–470, 2006.

[21] S. Mallat, “A wavelet tour of signal processing,” *Academic Press*, 1999.